JRC Scientific and Technical Reports

Analyzing Inter-relationships among Water, Governance, and Human Development variables in Developing Countries

Preliminary results on Africa for the year 2004

Celine Dondeynaz, Cesar Carmona-Moreno, Andrea Leone, Daoyi Chen

Ethiopia 2009- Water and Agriculture – Celine Dondeynaz

EUR 24605 EN

JRC EUROPEAN COMMISSION

ies Institute for Environment and Sustainability

The mission of the JRC-IES is to provide scientific-technical support to the European Union's policies for the protection and sustainable development of the European and global environment.

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information
Address: Celine Dondeynaz
via Fermi 2749, 21027 ISPRA, Itlay, (VA)
E-mail: celine.dondeynaz@jrc.ec.europa.eu
Tel.: +39 0332 78 5449
Fax: +39 0332 78 9960

http://ies.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC 59047
ISSN 1018-5593
ISBN-13: 978-92-79-18410-9
doi:10.2788/43519
Luxembourg: Publications Office of the European Union

Printed in Italy

Table of Content

Index of tables and figures

1. Introduction

The experience of the last 50 years of international cooperation[1] indicates that improving the understanding of the inter-relations among different related factors and variables linked with economic and human development is a fundamental topic of research.
This understanding is in fact an essential baseline in the design of development cooperation policies and strategies at national, regional and continental levels. In this way, looking at the Water sector in developing countries implies studying complex interactions between different environmental, socio-economic, governance and other human development factors.

This preliminary report (developed by the DG JRC, Institute for Environment and Sustainability – IES, Water Resources Management team of the MONDE action) briefly presents the first results analysing relationships among the different factors in the Water sector.

This report includes the description of the data sources, their selection, the description of methodologies used for verifying the data coherency and finally preliminary interpretations of results. In this first phase, our research focuses geographically on Africa.

The work was carried out through a series of steps which can be summarized as follows (details are given along this report):

a. A series of data related to human and economic development and to the water sector for the year 2004 for 52 countries in Africa have been collected. These data come from a series of international ”trusted” data providers (World Bank - WB, OECD, UNDP, UNEP, FAO, UN-HABITAT, etc…) but also from other national research institutions (Universities, research centres, etc...)
b. Data have been normalized and standard missing-imputation algorithms applied
c. Principal Component Analysis have been carried out to understand the relationships between the different variables and establish groups-clusters of similar behaviours
d. Data were linearly interpolated to get a first vision on the behaviour of the dataset. This interpolation cannot be interpreted as a modelling of the dataset, but only as a confirmation of the relationships between the variables deduced from the PCA analysis.
e. Finally, results have been interpreted
f. In a second phase, not included in this report, an analysis in deep the model will be performed from the relationships here described.

2. The approach of the research

In the early 90’s the international community adopted the integrated water resources management (IWRM) approach, considering water as a resource that should be holistically managed[2]. The effective management of the resource was acknowledged as central in order to provide sustainable WSS (Water and Sanitation Services) shifting away the main focus of the sector from the infrastructure development. In fact, the level of efficiency and development of water and sanitation services can be considered as the result of other factors. These includes the capacity of the country to manage the available water resource and its various uses at all scales (from local to national), as well as in general the socio-economic development of the country.

Looking at other fields, some research has been carried out following this integrated approach by performing comprehensive analyses of different variables or dimensions of a question. For instance,

[1] Easterly, W, 2001
[2] Principles laid down at the International Conference on Water and the Environment held in Dublin in January 1992
[2] Principles laid down at the International Conference on Water and the Environment held in Dublin in January 1992

Nicol Adler et al.[3] built a framework analysing human development index data, financial resources, and the Millennium Development Goals (MDG) target. The aim was to evaluate the efficiency of countries in progressing toward MDG's. Specific to the water sector, one study recently evaluated the relationships among the Official Development Aid, water and sanitation coverage and infant and child mortality[4].

Applying this cross analyse approach, this research aims at identifying the key elements explaining the various levels of access to Water supply and sanitation services observed. Using the standard MDG indicators[5] (the percentages of the population having access to improved water supply and sanitation), the objective is to map the variables impacting and/or influenced by the WSS level.

Relying on already measured variables through a wide range of indicators: environmental, social-economic, governance, indicators, we would like to answer the following main questions:

1. Are the different variables and data coherent enough to bring out relationships and spatial behaviours among them?

2. Can be established measurable protocols and can behaviour patterns be extrapolated in time and at other spatial scales?

3. Can data and patterns be integrated into a tool for better understanding these mechanisms?

The selection of variables and methodologies used to build the dataset is described in the next section. We restricted our analyses to the African countries for the year 2004, but an analysis across countries worldwide will be performed in the next step. The final objective of this research is to develop a methodological way of understanding this complex system of variables.

3. Methodology

3.1 Dataset construction

3.1.1 Logical framework

The data were chosen considering all variables that can result and can influence (double way relationship) the water supply and sanitation access levels. These variables have been clustered under four main areas or pillars (see Figure 1).

[3] N.Adler, E.Yahemsky, R. Taverdyan,(2009)

[4] M.J.Botting, E.O Porbeni, M.R Joffres,B.C Johnson, R.E.Black, E.J.Mills (2010)

[5]Millennium Development Goals indicators Provided by the United nations Statistic department for monitoring the progress toward the Target 3 of the Objective 7 about Water supply and sanitation services.

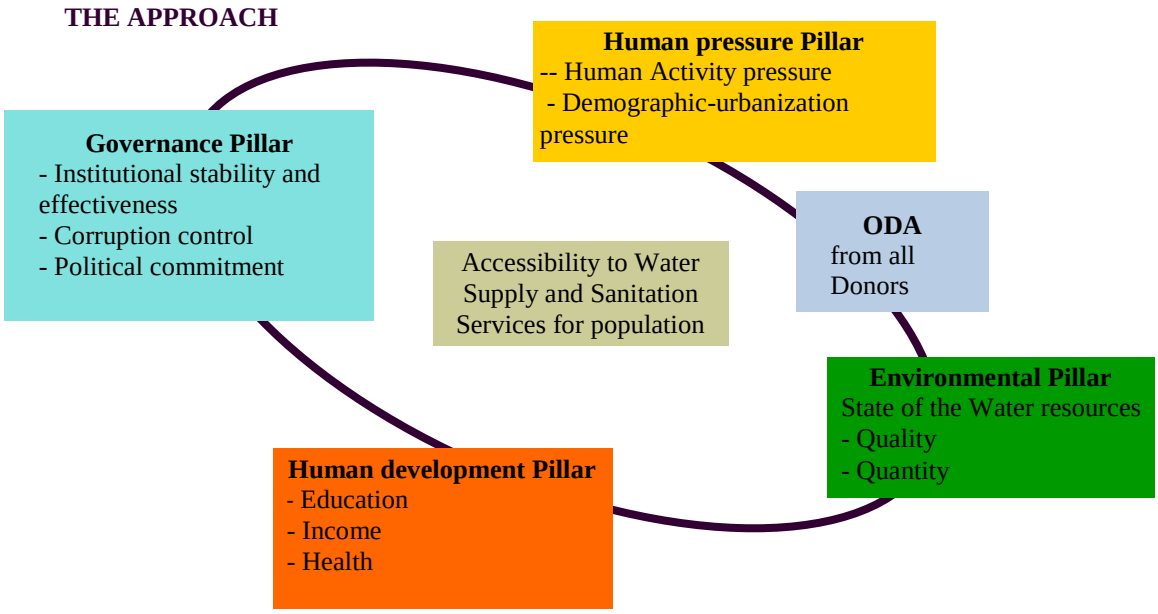


Figure 1: The Approach: factors-variables that influence and can be influenced by the level of development of WSS sector.

In addition to these four dimensions, as this research is oriented to developing countries, the **Official development Aid delivery** (ODA) in the water sector has been included in the database. This indicator represents the disbursed official aid provided to the developing countries. For detailed definitions of each variable selected please refer to Annexe 2.

3.1.2 The variables

Development indicators have been collected using data from official providers such as the World Bank, OECD, FAO, WHO, UN DESA, UNDP, UNSD, UN-HABITAT and research institutions such as Universities, NGO and Institutes (see Annexe 1).
The compatibility and consistency of this dataset in terms of geographical and temporal scales is a major constraint in the analysis process, e.g. indicators can be measured at a national scale or at the river basin scale. Finally, the national country scale was chosen as most of the data were given at that scale. The analysis of time series of data has not been considered in this first phase.
Year 2004 has been taken as time reference since the last release of the Joint Monitoring Programme report on WSS access level (JMP[6]) was based on data collected for that year.

The data collection covers countries worldwide.
132 indicators have been examined on the following main criteria:

- **Relevance**: the indicator has a potential role regarding the water supply or sanitation level of access,
- **Data availability**: The dataset has enough observations (less than 100 missing data over the 170 countries selected).
- **Reliability**: the indicator has been produced by trustfully providers using described methods.

After this first filter, 53 variables were finally selected and normalized. Complementary normalisation tests were performed for verifying the statistic stability of the variables.

[6] Joint Monitoring Programme http://www.wssinfo.org/datamining/tables.html

The behaviour coherency of the dataset of variables (the relationships between the variables and magnitudes of the values) were verified through a first run of Principal component Analysis (PCA). The following variables were removed because of too high correlation:

- Internal groundwater resources and internal surface water resources being respectively correlated at 0.848 and 0.779 with the Total water renewable resources.
- Population in Dryland being correlated at 0.817 with the Proportion of drylands in a country.
- The agriculture water demand couldn't be normalized, but as it's the most important component of the total water demand[7]. So in the PCA, the total water demand can be considered as a representation of the agricultural demand until a better normalization could be made.

The final database ends with a list of 48 variables.

NOTE: on data reliability: The data obtained are rather raw estimates (qualitative estimations) than exact quantitative values, mainly because of the nature of the indicators themselves and the context of developing countries.

3.1.3. Missing data treatment

The objective of the missing data treatment is rather to get realistic values for missing data than to have accurate values taking into account the nature of indicators we have collected.

With the characteristics of our dataset, we would need a multiple imputation method comparing country observations on several indicators in order to impute missing data without modifying the general statistic behaviour of the variables. The imputation method used in this study was "Expectation-Maximization Algorithm (EM)[8].

EM 's principles

Multiple imputation involves imputing *m* values for each missing cell in the data matrix and creating *m* "completed" data sets. Across these completed data sets, the observed values are the same, but the missing values are filled in with a distribution of imputations that reflect the uncertainty about the missing data.

EMB algorithm used combines the classic EM algorithm with a bootstrap approach to take draws from this in a second stage of the processing. For each draw, algorithm bootstraps the data to simulate estimation uncertainty and then run the EM algorithm[9] to find the mode of the posterior for the bootstrapped data.

Assumptions:

- The imputation model assumes that the complete data (that is, both observed and unobserved) are multivariate normal.
- The procedure uses the usual assumption in multiple imputations that the data are missing at random (MAR).

We completed our dataset step by step imputing missing data for variables starting from the ones with less missing data to the more incomplete ones. For getting better imputation, this process has been done on the worldwide dataset starting with 193 observations. The following country observations have been removed first because of the number of missing values and because being small states :

[7] Above 70% for developing world and respectively 39% and 32% for Nord America and Europe according to FAO Percentages on Agriculture withdrawal : http://www.unesco.org/water/wwap/wwdr/indicators/pdf/G5_Agriculture_water_withdrawals.pdf.

[8] Imputation have been made using Amelia II software provided by Honaker James, King Gary, Blackwell Matthew

[9] Honaker James, King Gary (2010)

French Guyana, Guadeloupe, Hong Kong, Macau, Martinique, Micronesia, the Dutch Antilles, Reunion, Virgin Islands, Serbia, Montenegro, Timor-Leste, Singapore, Aruba, Tonga, Puerto Rico, Trinidad and Tobago, Macedonia, Salomon Islands, New Caledonia, French Polynesia, Bahamas, Guam, and Barbados.

We encountered difficulties to set prior boundaries, but a manual rectification was performed in addition to a close check of the reasonability of the values imputed.
In case of few missing values (m<5) the Hot deck imputation method[10] has been used. It compares country observations on several indicators in order to impute missing data according to the "nearest neighbour" rule (see below).

Country	GDP-PPP per capita[1]	Industrial Withdrawal in %	Corruption Perception Index	BOD emission per capita[2]
Kyrgyzstan	1.721502	3	2.2	2.210746
Sudan	1.719801	1	2.2	NULL
Uzbekistan	1.712442	2	2.3	1.02716

[1] Gross Domestic Product- Purchase Power Parity in international US$

[2] Biological Oxygen Demand in kg/d

1.1<x <2.2
x ± 1.65

A) Hot deck imputation method

Looking at highly correlated variables to the BOD, Kyrgyzstan and Uzbekistan are close to those of Sudan. We impute the median distance between these two values.

Table 1: Example of Hot deck imputation Method

3.2 Statistical analysis methods

This preliminary analysis aims to test and validate the methodology and get an idea of the potential results before handling issues in normalization for the worldwide data. The dataset has been focused on Africa as the data normalisation could have been performed, enabling us to build multiple linear regression models as a first attempt. Indeed, the objective is to extend this approach to the full dataset using lessons learnt.

4. Preliminary results on Africa

Principal Component Analysis of all variables was performed for data from the 52 African Countries. Very small States (e.g St Helen) and French and Spanish territories were not taken into account because of data unavailability and/or reliability.

4.1 PCA performance

Composite indicators
GI Afr
HDI.
Water Poverty
ESI
HPI
WGI.W.A
WGI.PS.AV
WGI.GE
WGI.RQ.
WGI.RofL

Composite Indicators have been reprojected within the PCA to avoid bias due to the partial overlapping of their sub pillars with active variables.

The cumulated variability of the first three factors is equal to **50.386%** (Figure 2)
If we take into account the high heterogeneity and the nature of the variables, the three first Principal Components can be considered well enough for a first interpretation, but we will have to be careful when we interpret the maps as some information might be hidden in the next factors.

[10] Reilly Marie (1993)

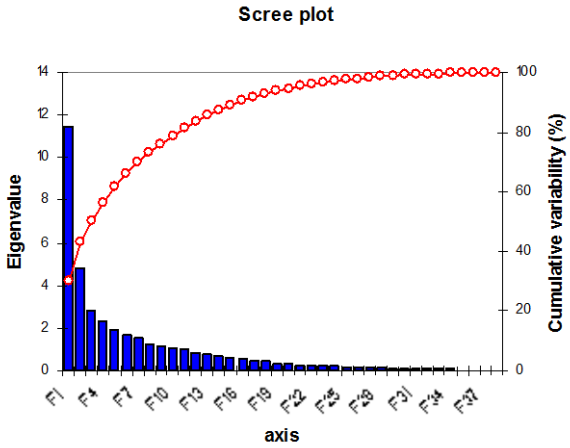


Figure 2: Variability distribution according to factors (F1 to F38)
The red curve indicates the level of cumulated variability that grows fast to reach 100%.

4.2 Analysis of the correlation between the variables

The PCA gives a clear and understandable picture of the relationships among the variables. At first sight, we observe several clusters of high-correlated variables (Figure 3) as described in the table below (See table 2).

Group 1: Country development indicators	**Group 2**: Poverty and demographic pressure
Life expectancy at birth (life – Expect) Income per capita (GDP-PPP) Ratio Girls to Boys at primary school (RatioGirls-boys) Gross global enrolment at school (School enrolment) Literacy rate for the Youth (literacy rate youth) Health expenditure (health expenditure) Level of access to Water Supply services (TOT-AIWS) Level of access to Sanitation services (TOT-AIS) Household connection proportion (water – house-connect) Proportion of irrigation in Agriculture (tot-irrigation) Biological Oxygen Demand (BOD) Urban population (Urban Pop) Water Poverty Index (Water poverty) Human development Index (HDI) Voice and accountability (GI – VA) Political Stability and Absence of violence (WGI – PS&AV) Government effectiveness (WGI – GE) Regulatory Quality (WGI – RQ) Rule of Law (WGI – RofL) Corruption Perception Index (CPI) Environmental governance (Environmental gov) Governance Index (GIAfr)	Growth Rural Population (GrowthRuralPop) Growth Urban Population (GrowthUrbanPop) Diarrhea rates in slums (% diarrhea in urban slums) Fertility Rates (FertilRates) Mortality rate of children under 5 years (Mortal_u5) Human Poverty Index (HPI) Children affected by Diarrhea (children with diarrhea) Poverty Rates Female economic activity rates (Female economic activity)
	Group 3: Agricultural pressure on water resources
	Proportion of Drylands in the country (X dryland) Participation to international environmental agreements (Particip in IEAg) Water Withdrawal total (Tot-WITH) Water Use Intensity in Agriculture(WaterUseInt.Agri)
	Group 4: water demand and water resources
	Water resources Withdrawal for domestic purpose (WITH-Dom) Water resources Withdrawal for domestic purpose (WITH-IND) Total Water Renewable Resources (TWRR) Amount of Precipitations (precipitat) Malaria incidence(Malaria) National Biodiversity Index(NBI) Water bodies surface (Water Bodies) Environmental sustainability index (ESI)

In blue the composite indexes

Table 2: Relationships and clusters of variables deduced from the first two PCA components.

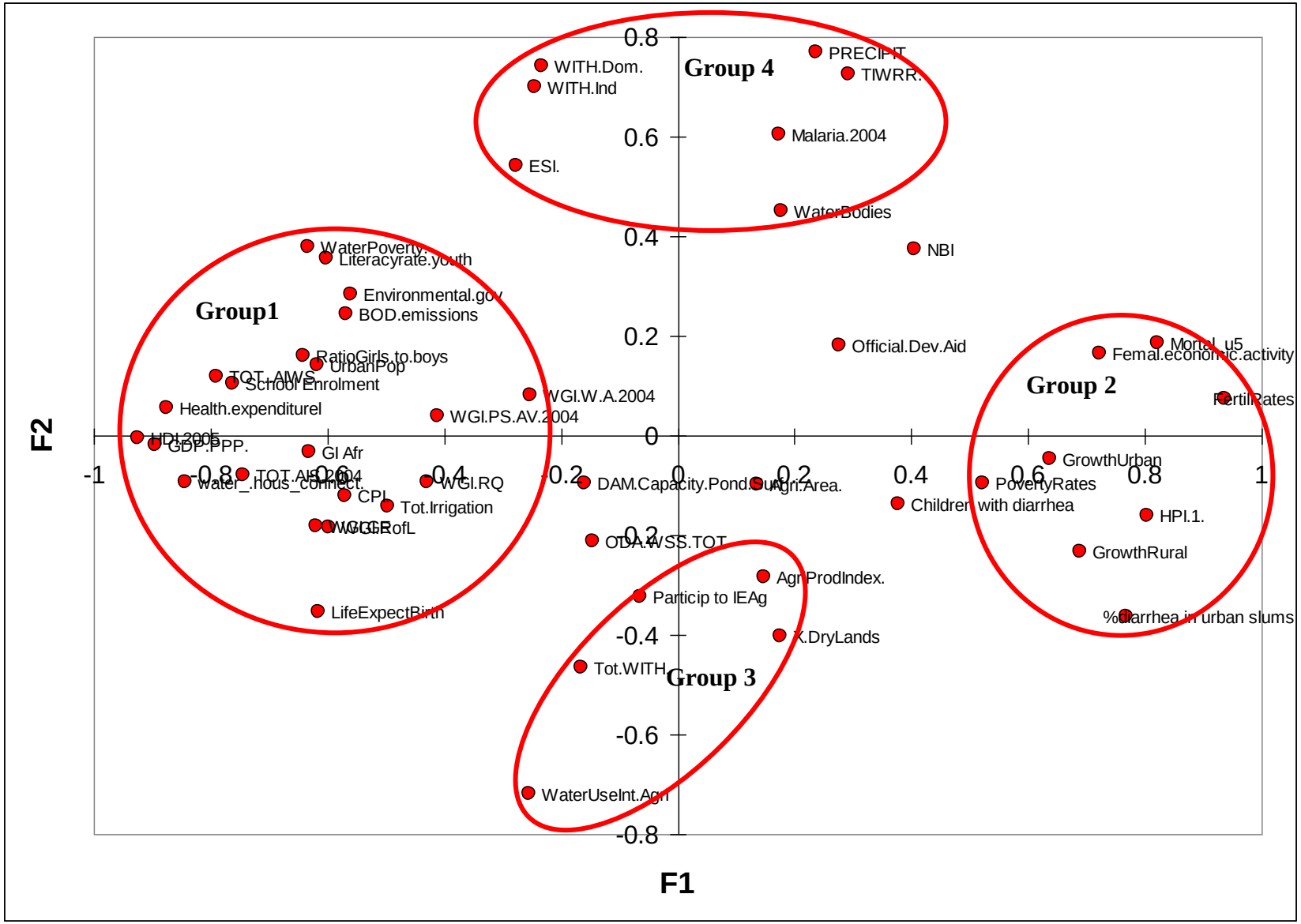


Figure 3: The first two PCA factors of variables, (accumulated variability equal to 43,02%)

Preliminary results

Note: these results are interpreted from the correlation matrix of the variables and the PCA above (Figure 3).

The general representation of the dataset on the different axes is coherent with what was expected: on the F1 axis, the positive development indicators (group1) are negatively correlated with poverty indicators (group 2). The F2 axis represents the indicators on water resources and water demand regarding in particular agricultural pressure (group 3 and 4).

Within the groups

- Governance indicators (Group 1) even calculated through various methods and data sources are coherent and highly correlated.
- The Environmental Sustainability Index is a very cross-cutting and complex index showing the capacity of a nation to manage its environment in a sustainable way. Therefore, as expected, it is highly correlated with many variables, but in particular the water resources availability, the level of urbanization (Domestic demand in water) and negatively with the agriculture pressure (water use intensity in agriculture).
- The Biological Oxygen Demand (BOD), an indicator of the water surface quality, is correlated with the urbanization level (Proportion of urban population, Urban pop and Urban slums). This indicator is highly correlated with the industrial and urban development of the country and cannot be taken as an indicative value of the water quality as such.

Relationships between groups

The following relationships can be deduced from the PCA above (figure 3)

- Governance indicators are positively correlated with the development of the country, the income (GDP-PPP, health expenditure, education rate), WSS access and negatively correlated with child mortality and fertility rates.
- The participation in international environmental agreements variable shows negative correlation with the water resources available (TWRR) and the amount of precipitations.
- The female economic participation belongs to the poverty group by being significantly correlated with the fertility rate and poverty rate. It's also negatively correlated to the level of the GDP [11], and the Human Development index (HDI) (see Annexe 3).

Variables near the centre of the graph

Their position in PCA do not allows us to identified significant relationships but call the following remarks:

- Regarding the dam capacity, no significant correlations can be observed with any other variables except the Agricultural area. Therefore, the capacity of countries to have reservoirs and dams does impact neither directly nor indirectly on the level of access to WSS.
- The agricultural surface is correlated with the National Biodiversity (0.414) in addition to the dam capacity. No conclusions can be reached because of bias due to the nature of the national biodiversity index.
- The national biodiversity index correlates with variables belonging to several groups (group 2 and 4 and for a fewer part group 1) contributing with the same approximate weight.
- The Children Diarrheal prevalence behaves mainly because of an inverse correlation with the variables in the group 1. Lower correlations are observed with the group 2.
- Financial flow either the global aid or the one specific to the WSS shows few significant correlations and is distributed among the 3 axes. However, the stability and the violence

[11] Mammen and Paxson (2000).

absence could be one of the criteria for Aid delivery. This need to be carefully verified while looking for other explicative variables that are missing.

5. Linear regression analysis

Linear regression methods are used to first refine the analysis of relationships between the different variables and second, explore how these variables could contribute to the explanation of the access level to water supply and sanitation services. Thus, this interpolation cannot be interpreted as a modelling of the dataset.

NOTE: This analysis does not show any evolution over time or evolution of country behaviour. We observe country values of several indicators around a specific year, 2004 and analyse this "picture" from the angle of variables.

Composite indicators need to be excluded from this preliminary analysis because they are too generic to explain specific behaviours of the variables to be analysed. Only Worldwide Governance Indicators (WGI) have been kept because expressing different aspects of the governance and thus not overlapping any other variables.

5.1 Sanitation service level

We start to include in the model the variables from the human development pillar because contributing the most to the explanation of the variability of the sanitation level variable (conclusion from preliminary tests). We include then the variables from the governance pillar, the environment and the human pressure pillar.

Parameters of the final model (Figure 4, Table 2)
Adjusted R^2 = 0.533 Complementary variables and data are needed to fully explain the variability of the Sanitation level. Two countries (Figure 4), Sao Tome and Principe and Gabon, are outside of the 95 % confidence interval of the model.

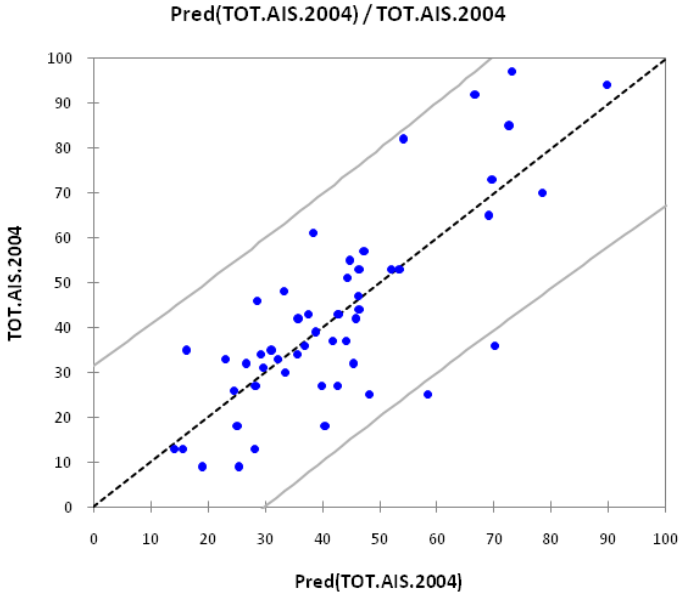
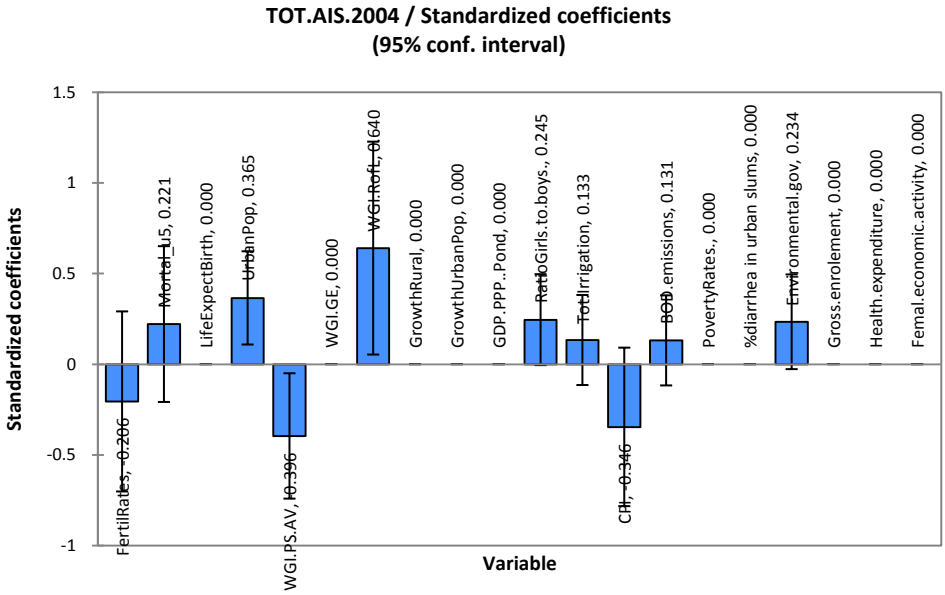


Figure 4: Observed values versus calculated values distribution
All observations should be within the 95% confidence interval (grey lines)



Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
FertilRates	-0.206	0.246	-0.836	0.408	-0.702	0.291
Children Mortal_u5	0.221	0.213	1.040	0.305	-0.208	0.651
UrbanPop	0.365	0.127	2.870	0.006	0.108	0.622
WGI.PS.AV	-0.396	0.171	-2.313	0.026	-0.741	-0.050
WGI.RofL	0.640	0.291	2.202	0.033	0.053	1.227
RatioGirls.to.boys.98.01	0.245	0.124	1.977	0.055	-0.005	0.496
Tot.Irrigation area	0.133	0.123	1.083	0.285	-0.115	0.381
CPI	-0.346	0.216	-1.599	0.118	-0.783	0.091
BOD.emissions	0.131	0.123	1.066	0.293	-0.117	0.379
Environmental.gov	0.234	0.129	1.808	0.078	-0.027	0.495

Table 3: Standards coefficients of the variables included in the model

Interpretation:

The sanitation level can be explained for 53% of its variability by:

- the governance aspects respectively the capacity to abide by rules and laws(WGI RofL), the level of stability without violence (WGI PAS.AV), and the corruption control (CPI) are main elements facilitating the expend of the Sanitation services. Secondary good environmental governance implies also a better level of sanitation services
- Sanitation is likely to be improved in urban areas.
- An unexpected point is the education of girl at primary level that contributes at a significant level while the general literacy rate of youth is excluded.
- The level of human development is linked to the sanitation access through the children morality-fertility rates. The irrigation capacity and BOD express the level of industrialisation or technical progress level of a country.

As explained before, the cases of Sao Tome and Principe and Gabon show different behaviours requiring more investigations.

5. 2 Water supply service level

As for the sanitation access level, we start to include in the model the variables from the human development pillar because contributing the most to the explanation of the variability of the sanitation level variable (conclusion from preliminary tests). We included then the variables from the governance pillar, the environment and the human pressure pillar.

Parameters of the most complete model (Figure 5, table 3)

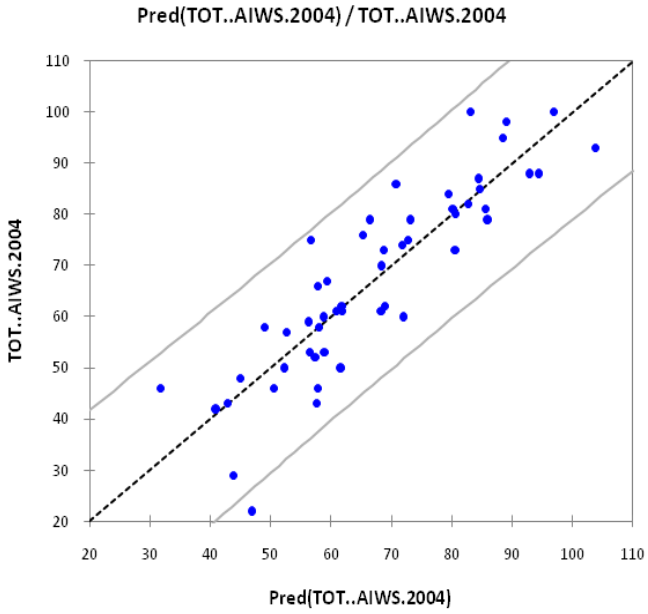
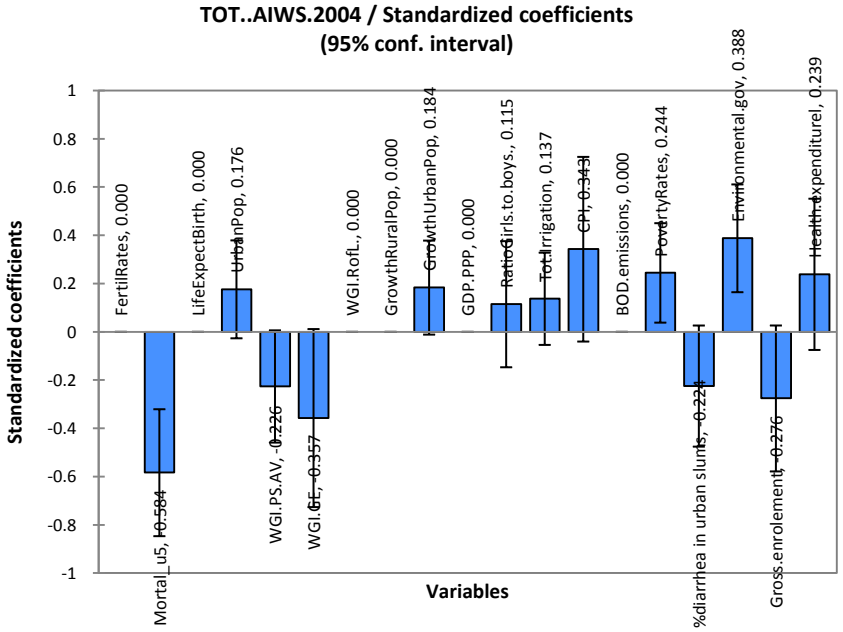


Figure 5: Observed values versus calculated values distribution
All observations should be within the 95% confidence interval (grey lines).

Adjusted R^2= 0.703. Complementary variables and data are needed to fully explain the variability of the Water supply level. Equatorial Guinea do not fit in to the 95% boundaries and thus need more investigations.



Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Children Mortal_u5	-0.584	0.130	-4.491	< 0.0001	-0.847	-0.320
UrbanPop	0.176	0.100	1.760	0.086	-0.026	0.379
WGI.PS.AV.	-0.226	0.115	-1.968	0.056	-0.459	0.006
WGI.GE	-0.357	0.182	-1.960	0.057	-0.726	0.012
GrowthUrbanPop	0.184	0.096	1.907	0.064	-0.011	0.378
RatioGirls.to.boys.	0.115	0.129	0.889	0.380	-0.147	0.376
Tot.Irrigation area	0.137	0.094	1.454	0.154	-0.054	0.328
CPI.2004	0.343	0.189	1.813	0.078	-0.040	0.725
PovertyRates.1987.2006.	0.244	0.102	2.401	0.021	0.038	0.451
%diarrhea in urban slums	-0.224	0.124	-1.811	0.078	-0.475	0.026
Environmental.gov	0.388	0.110	3.516	0.001	0.165	0.611
Gross.enrolment	-0.276	0.149	-1.846	0.073	-0.578	0.027
Health.expenditure	0.239	0.155	1.541	0.132	-0.075	0.552

Table 4: Standardized coefficients of the variables included in the model

Interpretation

The water supply services level can be explained for 70% of its variability by:

- Mortality of children under 5 is very linked to the Water supply services access.
- In term of governance, the environmental management capacity (environmental gov) as well as the stability, the implementation and service effectiveness, as well as corruption control are important.
- Living conditions meaning education level (Enrolment at school), efforts put in health improvement (health expenditure, diarrhea in urban slums), level of poverty (poverty rate) contribute to the explanation of the level of Water supply. Girls education is still include in the model as for sanitation even at a lower level.
- Finally the urbanisation process takes also a role.

6. Conclusions and way forward

Through this first analysis, this approach and methods showed their interest and capacity to quickly bringing to results. They also show the level of consistence and coherence of the dataset because of obvious relationships being verified. Besides, it also shows limits to be handled and thus we need to explore other ways for completing the explanation of the WSS level.

6.1 Main conclusions

With this methodology, we could explain around **53 %** of the variability of the sanitation variable and **70 %** for the water supply variable in Africa using 27 socio-economic and environmental variables. Based on the previous principal component analysis, we have performed regression with all variables represented in the first component for both sanitation and water supply. Usually merged into one issue, sanitation and water supply variables have diverse behaviours and so, ask for separated analyses.

Concerning sanitation:

The key elements are respectively the governance aspects, the urbanisation process and finally the human development level. Concerning the governance, special attention should be taken first to the

capacity of regulation of the country through police, courts, property right respect…(WGi RofL) then the control of the corruption (CPI), and finally the reinforcement of governance effectiveness in delivering services (WGI-GE). The urbanisation and living conditions complement the model but at a lower level. The cases of Sao Tome and Principe and Gabon have to be examined to determine reasons of diverse behaviour.

Concerning water supply
The water supply is very linked to the children mortality under 5 and the governance aspects. More precisely, it concerns the environmental governance, the capacity of being effective in delivering services (WGI-GE), finally, the political stability and absence of violence. Living conditions meaning education level (Enrolment at school), efforts put in health improvement (health expenditure, diarrhea in urban slums), level of poverty (poverty rate) contribute to the explanation of the level of Water supply. The urbanisation process comes at a lower level. The Equatorial Guinea has to be examined to determine reasons of its diverse behaviour.

Issues and Limits
Several remarks should be taken into consideration in reading the results:

- One should keep in mind that the data are reliable if considered as estimates.
- Regarding the methodology, the number of observations (52) is not sufficient with regards to the number of variables used (20). Two ways to increase the sample could be considered; on one hand performing the same analyses but on the whole geographical dataset or on the other hand, reducing the variables inserted into the linear regression model using group of variables. This remark is even more relevant as we demonstrate that additional variables are missing in order to reach the 100% of the variability being explained.
- Regarding the interpretation of results, we identify relationships between variables through the PCA but we didn't explain or give a particular direction to these relationships. It requires more analyses (maybe field work) in order to identify or validate causes or consequences explaining the WSS level.

6.2 Next steps and further research

Main lessons learnt and suggestions for a second phase:

- 70 % of the variability of the access to water supply services and around 53% for the access to sanitation have already been explained from the variables of the PCA first component. Other variables from the other PCA components should be included to complement and refine the analysis.
- The research was not defined to limit our investigations only to Africa and thus will continue and expand analyses at least at regional level, more preferably continental level. In addition, results on interrelations between WSS variables and human, economic and governance variables gain trust by having an important number of observations (values). The two above cited solutions will be initiated: expending the analyses on the whole geographical dataset while solving normalization issues and doing further analysis using groups of variables.
- The relationship between WSS and the education of girls at primary school asks for explanation on the role of girls into the WSS maybe also following a wider approach on gender issue.
- Governance indicators are composite indexes (compiled by the World Bank) that express six general aspects of the governance of a country. As governance is an essential element into the explanation of the access level of sanitation and even more into water supply services, it

would be useful to decompose these indices into more detailed indicators to be translated into identified and specific activities or actions.

- An interesting complement of this study will be the analysis of the data by country instead of variable. We could find out similar behaviour among countries and maybe define profile of countries having similar characteristics regarding socio- economic, environmental and governance aspects.
- These analyses were performed for the year 2004 and yet the possibility of having time series has not been examined in depth. However such analyses on variables and even on countries behaviours would be interesting to study trends over time and strengthen country profiles by adding progressive dimension.
- A specific study could be done to examine the differences as well as the behaviour of variables in rural and urban areas. Indeed, the JMP provides such a breakdown for water supply and sanitation services but still further investigations should be done to evaluate the availability of such data.

References

N.Adler, E.Yahemsky, R. Taverdyan (2009), *A framework to measure the relative socio –economic performance of developing countries*, Socio-economic planning Sciences, doi:10.1016/j.seeps.2009.08.001

D.Beguy (2009), *The impact of female employment on fertility in Dakar (Senegal) and Lome (Togo)*, Demographic research,20 (7), 97-127, Doi 10.4054/DemRes.2009.20.7.

M.J.Botting, E.O Porbeni, M.R Joffres, B.C Johnson, R.E.Black, E.J.Mills (2010), *Water and Sanitation infrastructures for health : the impact of foreign aid*, Globalization and Heath, 6:12, doi:10.1186/1744-8603-6-12

E.Boserup (1989), *Population, the status of women and rural development*, Population and development review, 15, Supplement: rural development and population: institutions and policy, 45-60.

W. Easterly, (2001), *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*, The MIT Press, ISBN 0-262-55042-3

D. C. Esty, M. Levy, T. Srebotnjak, and A. de Sherbinin (2005), *2005 Environmental Sustainability Index: Benchmarking National Environmental Stewardship*. NewHaven: Yale Center for Environmental Law & Policy, http://www.yale.edu/esi/

J Honaker, G King, M Blackwell, *Amelia II: A Program for Missing Data*, Version 1.2-17. http://gking.harvard.edu/amelia/

J Honaker, G King (2010), *What to Do about Missing Values in Time-Series Cross-Section Data*, American Journal of Political Science,54 (2), 561–581
http://gking.harvard.edu/files/pr.pdf

N J Horton, S R Lipsitz, *Multiple Imputation in practice: Comparison of Software Packages for regression models with missing variables*. The American Statistician, 55(3), 244-254.

K. Mammen, C. Paxson (2000), *Women's Work and economic Development,* Journal of Economic Perspectives, 14(4),141-164

M. Reilly (1993). *Data analysis using hot deck multiple imputation.* The Statistician, 42, 307-313.

R. I. Rotberg, and R.M. Gisselquist (2009), *Strengthening African Governance: Index of African Governance, Results and Rankings.* Cambridge, MA.
http://www.worldpeacefoundation.org/africangovernance.html

Online databases

AQUASTAT . a global information system on water and agriculture http://www.fao.org/nr/water/aquastat/main/index.stm

CPI, Transparency International all data are freely downloadable at
http://www.transparency.org/policy_research/surveys_indices/cpi/2004

EARTH TRENDS, the environmental information portal,
http://earthtrends.wri.org/searchable_db/index.php?action=select_theme&theme=1

GEO Portal, **Datasets** used by UNEP,
http://geodata.grid.unep.ch/#

JOINT MONITORING PROGRAMME
http://www.wssinfo.org/datamining/tables.html

OECD Portal
http://stats.oecd.org/qwids/

WORLD BANK
http://databank.worldbank.org/ddp/home.do#ranking

ANNEXE 1: Dataset structure

To be explained			Data provider
Access to Water Supply and Sanitation services		Access to Water Supply services Access to Sanitation services Household connection proportion	Joint Monitoring Programme Joint Monitoring Programme Joint Monitoring Programme
Pillars	Sub-pillars	Indicators	
Human development	Income	Human Development Index GDP- PPP Poverty Rate Human Poverty index Female economic activity rate	UNDP World Bank World Bank UNDP ILO
	Health	Malaria cases Proportion of children with diarrhoea Fertility Rate Mortality Rate for children under 5 Life expectancy at birth Health expenditure	WHO/UNICEF WHO/UNICEF WHO/UNICEF UNICEF UN DESA Population WHO
	Education	Ratio Girls to boys in preliminary school Literacy rate of Youth Gross enrolment at primary, secondary and tertiary school	World Bank World Bank UNESCO UNESCO
Environment state	Environmental general characteristics	Environmental Sustainability Index Water Poverty Index Percentage of dryland Surface of Water bodies Precipitations National Biodiversity Index	Yale and Columbia Universities, with the World Economic Forum, and JRC. FAO FAO FAO FAO Convention on Biological Diversity.
	Water quality	BOD release	FAO
	Water resources availability	total renewable water resources total internal water resources internal groundwater resources internal surface resources	FAO FAO FAO FAO
Human pressure	Water demand	Total Withdrawals Withdrawal by sectors of activities (domestic, industrial and agricultural demand)	FAO FAO
	Agricultural pressure	Agriculture area Agriculture production index Total surface in irrigation Water use intensity in Agriculture	FAO FAO FAO FAO
	Demographic pressure	Urban population Proportion of Population in dryland Urban Slum population	UN- HABITAT FAO UN- HABITAT
Governance	Governance efficiency	Voice and accountability Political Stability and Absence of violence Government effectiveness Regulatory Quality Rule of Law Control of corruption Corruption Perception Index Governance Index for Africa	World Bank World Bank World Bank World Bank World Bank World Bank Transparency international Harvard university
	Environmental concern	Environmental governance Participation to International Environmental agreements	World economic forum International conventions and agreements.*
Financial flow	Aid flow	Official Development Aid Disbursement for the water sector, breakdown by subsectors	OECD

*see in annexe 2 the list of agreements in the variable definition

ANNEXE 2: Variable definitions

Agricultural area
Agricultural area, this category is the sum of areas under a) arable land - land under temporary agricultural crops (multiple-cropped areas are counted only once), temporary meadows for mowing or pasture, land under market and kitchen gardens and land temporarily fallow (less than five years).

Agri Prod Ind, Agricultural Production Index
The FAO indices of agricultural production show the relative level of the aggregate volume of agricultural production for each year in comparison with the base period 1999-2001.

BOD release, Biological Oxygen Demand
Emissions of organic water pollutants are measured in terms of biochemical oxygen demand, which refers to the amount of oxygen that bacteria in water will consume in breaking down waste. It covers rivers, lakes and groundwaters.

Children with Diarrhoea
Proportion of children under 5 having or who have had diarrhoea within the two weeks before the survey.

CPI – Corruption Perception Index
CPI measures the perceived level of public-sector corruption in 180 countries and territories around the world. The CPI is a "survey of surveys", based on 13 different expert and business surveys.

Dyland Area
Dryland area as a percent of total area is the percent of the Earth's terrestrial area that falls within three of the world's six aridity zones—the arid, semi-arid, and dry sub-humid zones based based on the ratio of mean annual precipitation to mean annual potential evapotranspiration —as a percent of the Earth's total terrestrial area. This definition of drylands has been adopted by the United Nations Convention to Combat Desertification (CCD) to describe lands where problems with land degradation should be focused and where methods for attaining sustainable development should be promoted.

Population in Dryland
Proportion Estimate of population living in drylands

ESI, Environmental Sustainability Index
ESI benchmarks the ability of nations to protect the environment over the next several decades. It does so by integrating 76 data sets – tracking natural resource endowments, past and present pollution levels, environmental management efforts, and the capacity of a society to improve its environmental performance –into 21 indicators of environmental sustainability.
These indicators permit comparison across a range of issues that fall into the following five broad categories:
• Environmental Systems
• Reducing Environmental Stresses
• Reducing Human Vulnerability to Environmental Stresses
• Societal and Institutional Capacity to Respond to Environmental Challenges
• Global Stewardship

Female economic activity rates (2005)
This rate concerns women aged 15 and above and calculated on the basis of data on the economically active population (person looking for or having an occupation) and total population from ILO (International Labour Organization).
Generally, Students, retired people and persons not looking for an occupation are excluded.

Fertility Rates
Total fertility rate is an estimate of the number of children an average woman would have if current age-specific fertility rates remained constant during her reproductive years.

GDP-PPP, Gross Domestic Product – Purchasing Power Parity
Gross Domestic Product (Purchasing Power Parity) is gross domestic product converted to GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the U.S. dollar has in the United States. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current international dollars.

GI- Governance Index for Africa
This 2009 Index of African Governance measures the degree to which...five categories of political goods, i)safety and security ii) rule of law transparency and corruption iii) participation and human rights, iv)sustainable economic opportunity v) human development, are provided within Africa's fifty-three countries.

Girls to Boys Ratio in Primary Education Enrolment
This is a mesure of the attendance of girls at primary school.
The core at this level consists of education provided for children, the customary or legal age of entrance being not younger than five years or older than seven years. This level covers in principle six years of full-time schooling.

Gross enrolment ratio at school
It is calculated by expressing the number of students enrolled in primary, secondary and tertiary levels of education, regardless of age, as a percentage of the population of official school age for the three levels.
Health Expenditure (PPP- Capita)
Health Expenditure Per Capita (PPP; International $): The sum of public and private health expenditure (in PPP, International $) divided by population. Health expenditure includes the provision of health services, family planning activities, nutrition activities and emergency aid designated for heath, but excludes the provision of water and sanitation.

HDI, Human Development Index
The HDI (human development index) is a summary measure of human development. It measures the average achievements in a country in three basic dimensions of human development: - A long and healthy life, as measured by life expectancy at birth. - Knowledge, as measured by the adult literacy rate (with two-thirds weight) and the combined primary, secondary and tertiary gross enrolment ratio (with one-third weight). - A decent standard of living, as measured by GDP per capita (PPP US$).

HPI, Human Poverty Index
HPI is included in the HDI and takes into account three indicators qualifying deprivation :The first deprivation relates to survival: the likeliness of death at a relatively early age and is represented by the probability of not surviving to ages 40.The second dimension relates to knowledge: being excluded from the world of reading and communication and is measured by the percentage of adults who are illiterate. The third aspect relates to a decent standard of living, in particular, overall economic provisioning.

Internal renewable groundwater resources
Long-term annual average groundwater recharge, generated from precipitation within the boundaries of the country. Renewable groundwater resources of the country are computed either by estimating annual infiltration rate (in arid countries) or by computing river base flow (in humid countries).

Internal renewable surface water resources
Long-term average annual volume of surface water generated by direct runoff from endogenous precipitation (surface runoff).

Life expectancy at birth
Life expectancy at birth, both sexes is the average number of years that a newborn baby is expected to live if the age-specific mortality rates effective at the year of birth apply throughout his or her lifetime. The projections reported here assume medium fertility (the "medium-fertility assumption" of the United Nations Population Division).

Literacy rate of Youth
It is the percentage of people ages 15 -24 who can, with understanding, both read and write a short, simple statement related to their everyday life.

Malaria
Number of reported cases

NBI, National Biodiversity Index
This index is based on estimates of country richness and endemism in four terrestrial vertebrate classes and vascular plants; vertebrates and plants are ranked equally; index values range between 1.000 (maximum: Indonesia) and 0.000 (minimum: Greenland, not shown in table). The NBI includes some adjustment allowing for country size. Countries with land area less than 5,000 sq km are excluded. Overseas territories and dependencies are excluded from this column.
http://www.cbd.int/gbo1/annex.shtml

ODA, Official Development Aid
Aid as a percent of government expenditure is the amount of official development assistance (ODA) received by a country as a percentage of its central government expenditure

ODA-WSS Official Development Aid to the water sector
Total of ll donors disbursements of ODA towards all recipients related to Water supply and sanitation

Participation to international environmental agreements
It is calculated taking into account the participations to Framework Convention on Climate Change (UNFCCC), Vienna Convention on the Protection of the Ozone Layer, Convention on the Trade in Endangered Species (CITES), Basel Convention on the Transboundary Movement of Hazardous Waste and United nations.

Poverty Rate
National poverty rate is the percentage of a country's population living below the country's established national poverty line.

Total surface in irrigation
Area equipped to provide water (via irrigation) to the crops. It includes areas equipped for full and partial control irrigation, equipped lowland areas, pastures, and areas equipped for spate irrigation.

TIWRR, Total Internal Water Renewable Resources
Surface water produced internally: Long-term average annual volume of surface water generated by direct runoff from endogenous precipitation (surface runoff).

TWRR, Total Water Renewable Resources
This is the sum of the internal renewable surface water resources and the total external actual renewable surface water resources.

Under-five Mortality Rate
Probability of dying between birth and exact age 5. It is expressed as deaths per 1,000 births.

Urban Population -Rural Population
Total population residing in urban areas – in urban areas. Because of national differences in the characteristics that distinguish urban from rural areas, the distinction between urban and rural population is not amenable to a single definition that would be applicable to all countries. National definitions are most commonly based on size of locality. Population which is not urban is considered rural. The annual series of the urban population was obtained by systematically applying to the total population the percentage of urban population. The annual series of the percentage of the urban population have been derived through interpolation of the quinquennial series prepared by the UN (World Urbanization Prospects: The 2005 Revision).

Urban Slum population
Proportion of the urban population living in slums (A slum is a contiguous settlement where the inhabitants are characterized as having inadequate housing and basic services.)

Water Bodies Surface
It's the ratio of Water bodies regarding the total country surface. Water bodies are oceans, seas, lakes, reservoirs, and rivers. They can be either fresh or salt water bodies.

WPI, Water Poverty Index
WPI expresses an interdisciplinary measure which links household welfare with water availability and indicates the degree to which the water scarcity impacts on population .WPI has of five component indices: Resources, Access, Capacity, Use, and Environment.
The more this index is high, the lower is the water constraint is.

Water use intensity for agriculture
Water use intensity is the amount of water used in the agricultural sector per hectare of temporary and permanent cropland in the year specified. This indicator shows a country's dependence on irrigation for agricultural production

WGI-V&A, Voice and accountability
This index captures perceptions of the extent to which country'citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association and free media.

WGI PS&AV Political Stability and Absence of Violence
This index captures perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means including politically-motivated violence and terrorism.

WGI-GE, Government effectiveness
This index captures perceptions of the quality of the public services, the quality of the civil services, and the degree of its independence from political pressure, the quality of policy formulation and implementation and the credibility of the government's commitments to such policies.

WGI- RQ Regulatory Quality
This index captures perceptions of the ability of the government to formulate and implement sound policies and regulations to permit and promote private sector development

WGI-RofL Rule of Law
This index captures perceptions of the extent to which agents have confidence in and abide by the rule of the society and in particular the quality of the contract enforcement, property rights, the police and the courts as well as the likelihood of crime and violence.

WGI-CofC Control of Corruption
This index captures perceptions of the quality of public power is exercised for private gain, including both petty and grand forms of corruption, as well as capture of the states by elites and privates interests.
All WGI indexes come from many individuals sources on governance perceptions, been processed with the unobserved components model to construct aggregate measures.

TOT-WITH, Water Withdrawal Total
Annual gross quantity of water produced and used for agricultural, industrial and domestic purposes. It does not include other in situ-uses: energy, mining, recreation, navigation, fisheries and the environment, which are typically non consumptive uses of water. The typology of water use is independent from the source of water. Demands are covered by water productions: withdrawals from natural sources, fossil water abstraction (non-renewable production), non conventional water productions (reuse, desalination). The use of desalinated and treated wastewater is thus included. There are also referred to as non-conventional sources of water.

Total Water Use = Agricultural Water Use+ Domestic Water Use+ Industrial Water Use.

Some country, may use 'water withdrawal' when speaking about "water use" ; however it is not correct when non conventional waters are used (water withdrawal is then lower than water use).

WITH-Agr, Water withdrawal for agricultural purpose
It is an estimate of the proportion of water uses for agricultural purposes. Experience shows that the number of countries where agricultural water use is monitored with sufficient accuracy is limited. In most cases, gross irrigated areas are multiplied by an average unit water use to obtain an estimate of the country's water use in irrigation. Precipitation provides part of the water crops need to satisfy their transpiration requirements.

WITH-Dom, Water withdrawal for municipal purpose
Annual quantity of water use for domestic purposes. It is usually computed as the total amount of water supplied by public distribution networks, and usually includes the withdrawal by those industries connected to public networks.

WITH-Dom, Water withdrawal for industrial purpose
Annual quantity of water use by self-supplied industries not connected to any distribution network.

ANNEXE 3 Understanding the female economic participation variable

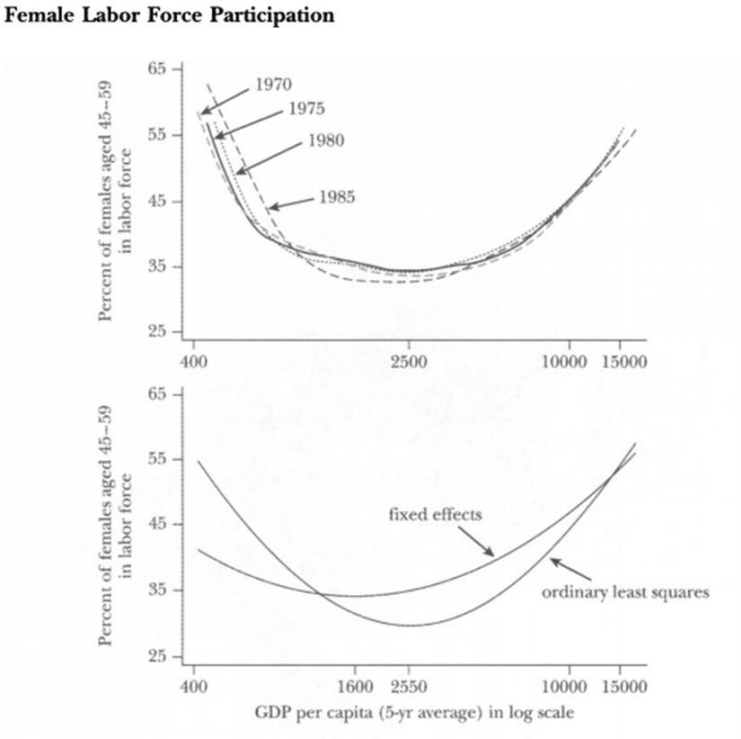


Figure 6: **Distribution shape of female participation in labour force** (for females aged 45-59 – but still true for other ages)

Mammen and Paxson (2000) explained that the female participation in labour force shows a U-Shape pattern when crossed with the income level per capita for the 1970-1985 period (see figure 6).
Female economic participation shows high rates for low income countries and decreases up to around 2550$ per capita. This exclusion of women from labour with the increase of income may have various causes: social norms and pressures, even legal constrains excluding women from work, economic additional costs if the women do not contribute to family work/farm, reduction of work opportunities for women.... The female participation starts increasing from 2550$.
Most African countries were under the 2500$ per capita for that period and still were under this threshold in 2004, except for countries in Magreb. Therefore, the decreasing trend observed on our data corresponds to the first part of the U-shape (see figure 7). This provides additional proofs on the coherency of the dataset.

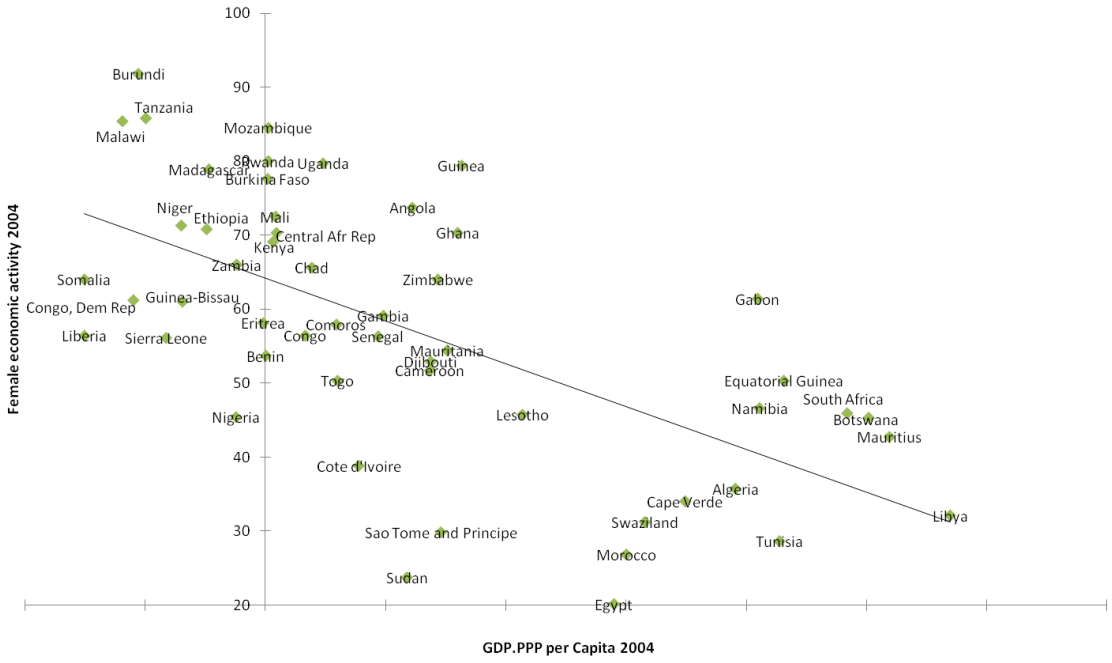


Figure 7: Female economic participation versus income per capita, distribution and trendline for 2004

The economic participation of women depends not only on the income but also on several other factors such as rural-urban context and fertility including social and cultural parameters that make more complex the explanation of this phenomenon (Ahn and Mira (1998), Boserup (1989), Beguy (2009)).

European Commission

EUR 24605 - Joint Research Centre – Institute for Environment and Sustainability
Title: Analysing Inter-relationships among Water, Governance, and Human Development variables in Developing Countries
Author(s): Celine Dondeynaz, Cesar Carmona-Moreno, Andrea Leone, Daoyi Chen
Luxembourg: Publications Office of the European Union
2010 – 30 pp. – 21 x 29.7 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN-13 : 978-92-79-18410-9
doi:10.2788/43519

Abstract

The experience of the last 50 years of international cooperation indicates that improving the understanding of the inter-relations among different variables linked with economic and human development can be an essential baseline in the design of development cooperation policies and strategies at national, regional and continental levels. In this way, understanding the Water sector in developing countries relies on complex interactions between different environmental, socio-economic, governance and other human development factors.

In this preliminary phase, this research focuses geographically on Africa. Data have been processed using the EM algorithm, hot deck imputation methods, logarithmic or square roots normalization to get a coherent dataset, baseline for performing statistical analyses.

Through this first analysis, this approach and methods showed their interest and capacity to quickly bringing to results. In fact, using Principal Component Analysis and linear regression analysis, we have explained a major part of behaviour of two variables: water services access level at **70%** and sanitation services access level at **53%**.

Main outputs are the ranking and weighting of variables according to their influence on the targeted variable. *For water supply access level*, the key elements are the governance aspects, in particular, the capacity of being effective in delivering services, the capacity of regulation through police, courts, property right respect….. and, finally, the control of the corruption in the country. *For sanitation access level*, the key elements are respectively education of girls far ahead the governance aspects. Concerning the latter, special attention should be taken first to the control of the corruption, then to reinforcement of governance effectiveness in delivering services and finally, the capacity of regulation of the country through police, courts, property right respect.

This study allows to move forward to more complex and detailed analysis having shown that data are enough coherent and reliable.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

LB-NA-24605-EN-C

ISBN 978-92-79-18410-9

Publications Office